

Chapter 6

Scoring

Open-response Questions and On-demand Writing

The 1998-99 Kentucky Commonwealth Accountability Testing System open-response questions and On-Demand Writing responses at grades 4, 5, 7, 8, 10, 11 and 12 required handscoring by Data Recognition Corporation personnel. While the processes of selecting and training scorers, reading and scoring papers, and monitoring scoring remained similar to those carried out in previous years for the KIRIS test, these procedures are described below in detail.

Staffing and Qualifications

Levels of staffing are listed in Table 6.1. The table also shows the number of scorers at each grade level who participated in a previous year's scoring (repeat scorers), as well as the number of training leaders. Table 6.2 shows education level and demographic information for scorers in the 1998-99 testing year.

Scorer Training

The training of scoring staff was completed in two phases: training of the scoring directors and team leaders, followed by training of the scorers.

The scoring directors for open-response questions in each content area met with the WestEd test developer responsible for a domain. The developer, as a facilitator of Kentucky's Content Advisory Committee (content specific), presented the Kentucky objectives, content guidelines, standards, and background information necessary to understand the objectives being measured. Each group also reviewed the framework of the scoring rubric and the language pertinent to the standard.

After this introduction, the combined group read hundreds of student responses and selected anchor papers — papers which typify each score point in the scoring rubric. Once the anchors were established, the scoring directors continued the preparation. They identified a second set of training papers, similar to the anchor set, which included current year examples of student responses that represented a range within each score point. Throughout this process, development staff was available to discuss concerns presented by the scoring directors and answer any questions that they might have. Before training sets were reproduced, the scoring directors met with the developers for a final review of the training materials.

TABLE 6.1 NUMBER OF SCORERS AND TRAINING LEADERS AT EACH GRADE						
Grade	1998-99			1999-2000		
	# Repeat Scorers (KIRIS)	Scorers	Training Leaders	# Repeat Scorers (KIRIS)	Scorers	Training Leaders
4/5	13	267	27			
7/8	96	322	30			
10/11/12	102	329	31			

TABLE 6.2 PROFILE OF SCORER QUALIFICATIONS AND DEMOGRAPHICS							
Background		Number of Scorers					
		1998-99			1999-2000 ¹		
		Grade 4/5	Grade 7/8	Grade 10/11 12	Grade 4/5	Grade 7/8	Grade 10/11 12
Education	Degrees beyond the Baccalaureate	41	68	73			
	Bachelor's Degree	173	209	213			
	Associate's Degree	20	25	24			
	Two-year college study or equivalent	33	20	19			
Demographics	Male	107	129	135			
	Female	160	193	194			
	Black	10	25	17			
	White	243	270	293			
	Other	14	27	19			

¹ 1999-2000 values will be available following the spring 2000 administration and analysis

WestEd development staff was present to observe the initial sessions when the scoring directors presented the standards to the scorers and was available to provide additional clarification when needed. The scoring director then completed the training independently. The scoring director and developer consulted as needed throughout scoring. At the end of the project, development staff and scoring directors met to share information about the process and to offer suggestions and comments for future improvement. Throughout the scoring period, scoring directors and team leaders read student responses as they checked the reliability of each scorer. As they read, scoring directors also selected recalibration papers that were reviewed with scorers to assure that drift from scoring standards did not occur.

Scorers for each content area were selected for their content expertise and were trained by the scoring directors. The scoring directors first presented background information and an explanation of the scoring rubric. The first set of training papers — the anchor training set — was used to clarify the language of the scoring rubric. Each score point was illustrated by several anchor papers. This set became the reference set used throughout scoring. Scorers were instructed to review the language of the rubric regularly as they read actual student responses.

The first training set was similar to the anchor set, but papers were in random rather than sequential order by score point. A second training set was designed to instruct scorers how to identify a range within each score point. After discussing the papers in each set, scorers were asked to assign scores independently to another set of papers. The scores were compared to those assigned by the scoring directors and item developers.

As a final qualifying step, scorers were instructed to score 15-20 responses on two separate qualifying sets. The scoring directors checked each reader's scores for accuracy. Those who achieved success on 80 percent of the papers in assigning the appropriate score point began to score actual papers. Those who needed further training worked with the scoring directors and team leaders until they were able to achieve the 80 percent rate of agreement required to qualify. All scorers hired for the project qualified in this process.

Consistency of Scoring

Scoring of open-response tests was monitored in two ways. The first was ongoing, as scoring directors and team leaders constantly moved from scorer to scorer, re-reading samples of each scorer's work. Each team leader read approximately one packet² per scorer each day.

The second monitoring procedure was a rescoring of two percent of the total reader's scores. Tables 6.3 and 6.4 document the percentage of exact agreement between scores assigned by separate scorers. These statistics indicate a high degree of consistency between scorers.

² A packet contains 15 student responses.

TABLE 6.3
INTER-RATER RELIABILITY IN SCORING OF OPEN-RESPONSE
QUESTIONS

Grade	Percentage of Exact Agreement					
	1998-99			1999-2000		
	4/5	7/8	10/11 12	4/5	7/8	10/11 12
Reading	81	80	86			
Math	83	84	87			
Science	76	81	68			
Social Studies	80	92	84			
Arts and Humanities	79	86	87			
Practical Living	82	79	80			
On Demand Writing	91	86	86			
Total	81.7	84	82.5			

TABLE 6.4 INTER-RATER RELIABILITY IN SCORING OF OPEN-RESPONSE QUESTIONS						
Percentage Within 1 Score Point						
Grade	1998-99			1999-2000		
	4/5	7/8	10/ 11/ 12	4/5	6/7	10/11/ 12
Reading	18	19	13			
Math	16	15	13			
Science	22	18	28			
Social Studies	18	8	16			
Arts and Humanities	20	14	13			
Practical Living	18	20	20			
On Demand Writing	9	14	14			
Total	17.2	15.4	16.7			

TABLE 6.5 INTER-RATER RELIABILITY IN SCORING OF OPEN-RESPONSE QUESTIONS						
Grade	Percentage of Non-adjacent Score Point					
	1998-99			1999-2000		
	4/5	7/8	10/ 11/ 12	4/5	7/8	10/11/ 12
Reading	1	1	1			
Math	1	1	0			
Science	2	1	4			
Social Studies	2	0	0			
Arts and Humanities	1	0	0			
Practical Living	0	1	0			
Writing	0	0	0			
Total	1	.57	.71			